The Nevada Adequate Yearly Progress Technical Manual

June 11, 2004

Prepared by:

Paul M. La Marca, Ph.D.
Assistant Deputy Superintendent
Nevada Department of Education

For additional copies, please visit our website at www.doe.nv.gov or contact the APAC office at (775) 687-9186. Please submit questions to Paul La Marca at plamarca@doe.nv.gov or at (775) 687-9180.

The Nevada Adequate Yearly Progress Determination Process

On January 9th, 2002 President Bush signed into law the Federal No Child Left Behind Act (NCLB—HR 1). This reauthorization of the federal Elementary and Secondary Education Act and its sweeping reforms has impact on every state including Nevada. In response to the new federal law, the Nevada Legislature significantly revised its own accountability statutes through passage of Senate Bill 1 in the 19th Special Session (June, 2003). This was a necessity given significant differences between existing state and federal statute. At the heart of both the federal and revised state statutes is a conservative school, school district, and state accountability model working under the auspice of guaranteeing all students the opportunity for and access to a challenging and meaningful educational experience. Toward this end and on an annual basis, schools, school districts and the state as a whole are judged against a set of adequate yearly progress (AYP) criteria. The judgment of success is based largely on performance on assessments aligned to state content standards administered on an annual basis, and by attending specifically to the performance of disparate subgroups of students.

The date of passage of Senate Bill 1 is significant in its own right. A brief timeline may help to dispel some of the mystery that has hampered the implementation of the federal act in Nevada.

January 9th, 2002 – President signs the NCLB Act
January 9th, 2002 – States are required to immediately begin implementing NCLB mandates including AYP determination of schools/school districts based on 2002-03 test performance
August, 2002 – Federal government releases regulations governing NCLB assessment requirements
December, 2002 – Federal government releases regulations governing the majority of NCLB accountability requirements
January 31, 2003 – States required to submit draft accountability plan to federal government for peer review
February 1, 2003 – Nevada 2003 Legislative Session begins
March 16, 2003 – Nevada draft accountability plan peer reviewed by federal government
May 31, 2003 – States must submit final accountability plan to federal government
June, 2003 – 19th Special Session of the Legislature begins. Of key interest is Senate Bill 1, previously known as Senate Bill 191 which failed to make it out of the 2003 Session
July, 2003 – Formal approval of Nevada accountability plan by federal government

To summarize, states have been required to implement NCLB reforms with no real phase in period. In Nevada this has meant that school administrators were administering tests to students without knowledge that the tests would be used in determining the relative success/failure of their schools within an AYP framework. Although working documents regarding the Nevada accountability plan and its implementation were available to school districts throughout the 2002-03 school year, firm and codified policy was lacking. The problems associated with a lack

of formal documentation regarding the new system were exacerbated the conflict between existing state statute and regulation.

It is difficult to gauge how the credibility of the new accountability model has suffered as a result of inadequate communication but it is the sincere intent of the Nevada Department of Education to overcome these obstacles. The following is an overview of the Nevada AYP Determination Process. An attempt has been made to balance the need for detail with the need to clearly communicate critical elements of the plan. Unfortunately, this does not constitute a culminating exposition on this matter. The Nevada AYP model is likely to see significant change in years to come and every attempt will be made to communicate these changes clearly and on a timely basis. As a result the following should be considered a living document. This manual will focus on the following elements:

- The Assessment system
  - Emerging system
  - Test alignment
  - SCAAN/LEP testing
- Determining AYP
  - AYP profiles
  - Step by Step—part rate, status, safe harbor, OI
  - Operational definitions
  - Key concerns
- Timelines
  - Assessment
  - AYP
  - Reporting
  - School improvement
- Reporting
  - Reporting N-size
  - Contents of state, district, school report cards
  - What other accountability reports are required?
- Special schools and circumstances

It is important to note that the AYP determination process applies equally to schools, school districts, and the state. Although determinations being made relative to schools will guide the presentation of information included in this bulletin, the same general rules will be applied to school districts and the state. Note will be made when differences in rules do exist.

The Assessment System

The foundation for the NCLB accountability system is made up of state content and achievement standards and large scale assessments designed to measure the standards. NCLB expanded previous requirements regarding the development of state standards in English Language Arts and Math by requiring states to develop content and achievement standards in science.

Similarly, NCLB expanded assessment requirements from previous legislation by requiring states to develop and implement tests in grades 3 through 8 and in at least one grade at the high school level in English Language Arts and Math. These tests must be fully implemented by the 2005-06 school year. Additionally, by 2007-08 states must develop and implement science tests to be administered in at least one grade in three separate grade ranges (3-5, 6-8, 9-12).

A couple of other assessment related statutory requirements are pertinent to the emerging Nevada assessment system. Assessments developed to fulfill the above requirement and for use in the AYP process must directly align to state content standards. This may seem in retrospect as an obvious requirement given the stature and role of state content standards. However, this translated in a near whole scale exclusion of the use of norm-referenced tests by the federal government and the exclusive use of criterion-referenced tests. This required significant changes for states such as Nevada that had in the past relied so heavily on the use of norm-referenced tests. Through the regulation process, the federal government softened its position on the use of norm-referenced tests allowing what has been referred to as augmented norm-referenced tests placing the burden on states to substantiate the alignment between the customized norm-referenced tests and a state's content standards.

The alignment condition also has a direct relationship to the federal requirement that states employ multiple measures in assessing student achievement. The use of multiple measures is expected to enhance both the reliability and validity of the measurement process. It does this by enabling an expansion of the content assessed and allowing the convergence of findings from disparate measurement methods. The relationship to alignment is evidenced when we draw a distinction between breadth of content coverage and depth of content coverage.

Breadth of coverage is achieved by sampling widely and representatively from the stated or prescribed content. Depth coverage refers to the degree by which the cognitive demands elicited by the assessment tasks match the cognitive demands prescribed by the content standards. For example, we expect students to be able to write persuasive and informational essays that clearly articulate ideas, are well organized, employ appropriate voice, and are conventionally sound. For measurement purposes, we have available to us multiple choice questions that correspond to idea articulation, organization, voice, and conventions. Student responses to these questions may tell us something about writing ability. However, there may be alternate methods available that provide a more direct measurement of writing skill and that more directly tap the cognitive demands prescribed by the standards. In other words, having students respond openly to a writing prompt may entail a very different cognitive process than having students respond to an editing activity.

Understanding this distinction and the federal requirement, state assessments used in the AYP process will involve multiple question formats. Criterion-referenced tests will include both multiple choice and constructed response items and the state will continue to employ performance based writing activities.

Additionally, tests used in the AYP process must align with state achievement standards. In essence, this means that the tests must enable a distinction to be drawn between students that are basic (below proficient), proficient, and advanced. The federal government allowed flexibility in

the labeling of achievement levels and the number of achievement levels so long as the above distinctions were met. In Nevada four achievement levels are used with two levels identifying performance that is below meeting standard or proficient.

The key point to be made is that each assessment used in the AYP process must yield information that can be categorized using the achievement level distinctions. As will be discussed in more detail, this similarity among all included assessments allows the variety of assessments to be combined when making AYP determinations.

Crosswalk of Nevada and Federal Achievement Level Categories

Nevada Achievement Levels	Federal Achievement Levels
Developing/Emergent	
Approaching Standard	Basic
Meets Standard	Proficient
Exceeds Standard	Advanced

States are given the responsibility to determine what level of performance (cut-scores) on its state specific tests is indicative of proficiency or meeting the state's content standards or expectations for student knowledge and skill attainment. States must employ objective methodologies that rely heavily on the professional judgment of educators in making these decisions.

To comply with the NCLB Act and the more recent state statutory changes, the emerging state assessment system is planned as follows:

	2002-2003	2003-2004	2004-2005	2005-2006
Grade 3	CRT—Reading, Math	CRT—Reading, Math	CRT—Reading, Math	CRT—Reading, Math
Grade 4	NRT—ELA, Math, SCI Perf—Writing	NRT—ELA, Math, SCI Perf—Writing	NRT—ELA, Math, SCI Perf—Writing	NRT—ELA, Math, SCI CRT—Reading, Math Perf—Writing
Grade 5	CRT—Reading, Math	CRT—Reading, Math, SCI	CRT—Reading, Math, SCI	CRT—Reading, Math, SCI
Grade 6				CRT—Reading, Math
Grade 7	NRT—ELA, Math, SCI	NRT—ELA, Math, SCI	NRT—ELA, Math, SCI	NRT—ELA, Math, SCI CRT—Reading, Math
Grade 8	Perf—Writing	CRT—Reading, Math, SCI Perf—Writing	CRT—Reading, Math, SCI Perf--Writing	CRT—Reading, Math, SCI Perf--Writing
High School	NRT—ELA, Math, SCI HSPE—ELA, Math, SCI	NRT—ELA, Math, SCI HSPE—ELA, Math	NRT—ELA, Math, SCI HSPE—ELA, Math	NRT—ELA, Math, SCI HSPE—ELA, Math

CRT = Criterion-referenced tests; NRT = Norm-Referenced Tests; Perf = Performance-Based Test; HSPE = High School Proficiency Examinations
Shaded areas = tests used in the AYP determination process

Based on objective methods, the following table provides a summary of test cut scores that correspond to the achievement levels used in the AYP determination process.

Grade	Test	Subject	Developing/ Emergent	Approaches Standard	Meets Standard	Exceeds Standard
Grade 3	CRT	Reading	100-199	200-299	300-355	356-500
		Math	100-199	200-299	300-354	355-500
Grade 4	Performance	Writing	0-7.5	8-11.5	12-15.5	16-20
Grade 5	CRT	Reading	100-199	200-299	300-384	385-500
		Math	100-199	200-299	300-380	381-500
Grade 8	CRT	Reading	100-199	200-299	300	385-500[a]
		Math	100-199	200-299	300	385-500[a]
	Performance	Writing	0-7.5	8-11.5	12-15.5	16-20
Grades 10 & 11	HSPE	Reading	TBD	TBD	251	TBD
		Math	TBD	TBD	290	TBD
		Writing	TBD	TBD	7	TBD

[a] Exceeds cut points for 8th grade tests are estimated

The assessments included in the above table do not constitute an exhaustive set of state assessments. States, based on additional federal mandates, must employ the use of alternate assessments for students with severe cognitive deficits. It is through the student's individualized educational program that decisions are made regarding the assessments to be administered to the student. Although states must employ an alternate assessment for this subgroup, states are limited in the number and percentage of students that can demonstrate proficiency through an alternate means of assessment. At this juncture, the federal government has capped the percentage of students that can be deemed as proficient through use of alternate special education assessments at 1% of the total student population. We will return to this issue later.

States are also required to annually assess the language acquisition and proficiency of students that have limited English proficiency. These students must be tracked as they make progress relative to English language acquisition and after they exit from program participation status. LEP students, with the exception of immigrants having been in the U.S. for less than a full year, must still be included in the accountability system and assessed relative to the same state content requirements.

In compliance with the NCLB Act, the Nevada accountability plan allows school districts to opt for use of alternate content based assessments for limited English proficient students who have been in the United States for less than a 3-year period. On a cases-by-case basis, districts may choose to extend the use of alternate assessments for two additional years. The use of alternate assessments must be approved by the Nevada Department of Education prior to use. Approval is based on a review of technical characteristics of the assessment. In short, the alternate assessments must be aligned to the state's content and achievement standards and yield reliable and interpretable performance information.

Section Summary

As implied above, the requirements for standards and assessments within the NCLB Act lay the foundation for the AYP determination process. Much debate has ensued regarding the pros and cons associated with the weight being placed on large-scale assessments. It can be argued persuasively that only through annual ongoing assessment can we understand the accomplishments and needs of students. By contrast, arguably the amount of time being dedicated to assessment takes away from precious time needed to provide students with proper instruction.

Without taking a side, an important issue must be raised. Although the Act calls for annual assessment in grades 3 through 8, the Act does not intend for the assessments to be used to track student progress longitudinally. This seems counterintuitive given the annual assessment requirement.

It seems obvious that the spirit or intent of the assessment system is to identify strengths and weaknesses of students early (i.e. grade 3) and to then individualize instruction to better ensure that each student eventually realizes state academic expectations (i.e. makes gains annually with a culminating demonstration of proficiency or advanced performance). However, as will be described in detail, on an annual basis the AYP determination process is solely interested in the percentage of students within a school (school district and state) that demonstrate proficiency. As a result, some schools will be identified as meeting AYP requirements and others as failing to meet AYP requirements.

After being identified as failing to meet AYP, more students or a greater percentage of students will have to be proficient in the future to meet the AYP criteria. But the AYP determination process makes no real attempt to quantify growth or reward it. A school can show no change or actually show a decline in the percentage of proficient students and still be recognized as adequate while another school may demonstrate a significant positive change in the percentage of proficient students but still be recognized as less than adequate.

Regardless of personal feelings regarding the assessment debates, it must be recognized that large scale assessments are a key ingredient to the federal educational reform movement. Given the stakes associated with negative AYP classifications, they will undoubtedly drive instructional practice. It is our charge to develop appropriate and meaningful assessments, ones clearly aligned to our state expectations for skill and knowledge attainment, and to build a system of support that assists schools in the proper use of state assessments and interpretation of results.

Determining AYP

On an annual basis, schools, school districts, and the state as a whole must be judged regarding the adequacy of their progress. As discussed above, the inclusion of the word progress in the annual judgment process lacks precision. We are really talking about static performance and not progress. Regardless, there is some intuitive appeal to the AYP process. The primary determinants of adequacy are objective assessments and a basic interpretation of assessment

results. The basic question is if a sufficient percentage of students are demonstrating satisfactory knowledge and skill attainment. If yes, good; if no, bad. But as will be revealed, the determination process is quite complex. There are a multitude of factors to consider before that final judgment can be made.

To assist the reader, the following discussion of the AYP determination process will use the school as the level of analysis. As noted above, school districts and the state as a whole must also be judged. By and large, the same rules used to judge schools are used when judging school districts and the state. Attention will be drawn to those limited instances where differences exist.

Subject Area Achievement Indicators

AYP is determined separately for English Language Arts (ELA) and Math. At this time, the federal mandate does not require the inclusion of science results. For each subject, the state must establish annual goals indicating the minimum percentage of students that must score at or above the meets standard level of achievement on the Nevada AYP tests. This is often referred to as the percent at/above cut or the PAC or the status comparison. If a group does not meet the status goal but demonstrates a decrease in the percentage of non-proficient students of at least 10% from the previous school year and meets the other indicator criteria, the group can be judged as meeting the AYP achievement indicator. This is referred to as the Safe Harbor provision.

Participation Indicators

Schools are required to have at least 95% of all students participate on the state AYP tests to meet the AYP requirements. Participation rates on English language arts and mathematics tests are considered separately.

"Other" Indicators

In addition to subject area proficiency and test participation, schools must be judged with respect to at least one "other" indicator. At the high school level, the NCLB Act requires that graduation rate be used. The Act gives states flexibility in the use of other indicators at the elementary and middle school levels. State statute now requires that elementary and middle schools in Nevada be judged relative to average daily student attendance.

Similar to the achievement comparisons, school performance on the other indicator is compared against an annual statewide goal. If the static goal is not met but the school demonstrates improvement in comparison to the previous year, the school can be judged as meeting AYP.

Subgroups

ELA participation, ELA achievement, math participation, math achievement, and, in certain instances (i.e. safe harbor analyses if necessary), other indicator performance are judged separately for 9 separate subgroups. These include the school as a whole, five major race/ethnic subgroups (American Indian, Asian/Pacific Islander, Hispanic, African American, and white), students with disabilities, students with limited English proficiency, and students who are economically disadvantaged (economic disadvantage is predicated on eligibility for free and/or reduced lunch). It is important to note that individual students are counted multiple times when we consider subgroup analyses. For example, the majority of students in Nevada with limited English proficiency are also classified as Hispanic. Many of these same children are eligible for

free and/or reduced price lunch and of course they are counted as part of the whole school as well.

In brief, a school and each of its identifiable subgroups have several hurdles to jump in order for the school as a whole to be deemed as demonstrating adequate yearly progress. First, a minimum of 95% of the students from each subgroup must participate on the state tests. Second, each subgroup must meet the achievement indicator (i.e. statewide status goal or the safe harbor provision) in each subject area. Third, the school as a whole must meet the other indicator criterion. This also applies to any subgroup within a school that fails status but meets the safe harbor provision.

It becomes clear that to make AYP, the state and school districts must conduct many comparisons for each school and the school must pass each satisfactorily. By contrast, a failure with respect to a single comparison may lead to a negative AYP classification. Presented below in tabular form is a summary profile of the basic comparisons that must be made when determining AYP.

Population	ELA Participation	ELA Achievement		Math Participation	Math Achievement		Other Indicator
		ELA Status	ELA Safe Harbor		Math Status	Math Safe Harbor	
School	Yes	Yes	-	No	Yes	-	Yes
American.Indian/ Alaskan Native	Yes	***	***	Yes	***	***	
Asian/ Pacific Islander	Yes	***	***	Yes	***	***	
Hispanic	Yes	No	No	Yes	Yes	-	
Black	No	No	No	No	No	No	
White	No	Yes	-	No	Yes	-	
IEP	Yes	No	No	Yes	No	No	
LEP	Yes	No	Yes	Yes	No	Yes	
FRL	Yes	No	No	Yes	No	No	

The number of comparisons depends in large part on the number of identifiable subgroups within a school and the number of times the safe harbor provision must be used to judge subgroups. In the next section we will describe a number of key methodological issues and operational definitions for the various indicators.

Operational Definitions and Key Methodological Considerations

One of the initial steps in organizing information in preparation for the AYP analyses is to identify the population of students to be included. No Child Left Behind allows achievement indicators reflecting school and disaggregated group performance to be based only on students who have been enrolled for a full academic year. By contrast, test participation and "other" indicator performance do not include this filter in defining the eligible student population. Instead, all students must be included. Hence, we need a definition of full-academic year and we need to be able to apply it to filter the included population at appropriate times.

- A student is considered to be enrolled at a particular school for a full academic year if he or she has been continuously enrolled from the official count day of students, occurring in early fall, until and during the specified test window, which occurs in mid-Spring.[1]

A second key issue is the requirement that states build reliable and valid systems of determinations. Much debate has ensued regarding how or what approaches are best suited to support the reliability of the system. There are two assumptions that predicate some of the choices made in Nevada. First, the more individual comparisons made to profile a school, the greater probability that a school will be identified as failing AYP. Second, the greater the proportion of students within a school included in the assessment system and, hence, AYP determination, the greater the probability that the final determination is representative of the school.

Data Aggregation

Following this logic, to the benefit of schools, data is aggregated across grades when making AYP determinations. For example, the most typical grade configuration among elementary schools in Nevada is a K-5 structure. As noted above, assessments are administered in grades 3, 4, and 5. Therefore, all assessment related indicators reflect an aggregation (adding) of students across grades 3, 4, and 5.

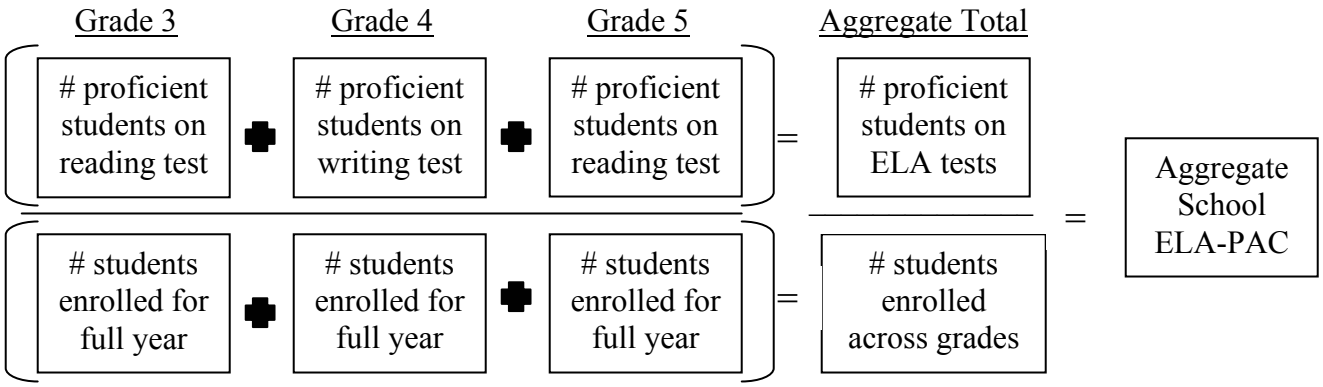
Looking back at the assumptions, aggregating across the 3 grades means far fewer school level comparisons (in contrast to grade by grade comparisons) and a more reflective population with some control over the effect one particular cohort of students may have on a school (i.e. a poor performing 3rd grade cohort may be offset by higher performing 4th and 5th grade cohorts).

Data aggregation is not accomplished easily. It is most challenging when as part of the aggregation, combinations of different tests must be considered. For example, for the 2003-04 determinations, data will have to be aggregated across grades 3, 4, and 5 to estimate ELA performance. This will include reading tests in grades 3 and 5 and a writing test in grade 4. In addition, some IEP students will take the SCAAN alternative for ELA and it is possible that school districts may use a content alternative for LEP students.

This means that aggregation will include 3 grades and at least 3 different tests. The greater challenge is combining information among the disparate tests. The challenge is met by combining the tests using achievement level scores. As noted above, each test used in the system must align to state content and achievement standards. The tests align to the achievement standards by yielding achievement level scores (i.e. meets standard). These scores provide a general statement regarding a student's overall performance relative to Nevada standards.

[1] Note that the same rules apply to school districts. A student is considered in the achievement-based analysis for school districts if he or she has been in the school district for a full-academic year. The difference is that a student could be included in a school district analysis even if they had attended two or more schools during the school year, so long as the different schools are all part of the same school district. For the state AYP analysis, all students, regardless of years in school/years in district are included in the analysis.

Although the reading test at grade 3 taps different content standards than the 4th grade writing test, each provides an estimate of student proficiency relative to ELA standards more generally. From this logic, we combine the numbers of proficient students for each test and divide that number by the aggregated grade level enrollment counts. Graphically the aggregation looks as follows:



In short, data aggregation is one of the measures employed to meet the NCLB requirement of a reliable and valid system. Through aggregation we can minimize the number of comparisons a school is subjected to, provide a better proportional representation of the school, and, hence, increase the stability or reliability of the data used to make comparisons. A second important measure taken in Nevada with the goal of reliable decision making is discussed next.

N-Size, Confidence Intervals, and Systems Reliability & Validity

As noted above, it is assumed that greater proportional representation of a school enhances the reliability of school performance estimates. This should not be taken to imply that some specific number of students is required to reliably represent a school or subgroup within a school.

NCLB requires that states employ reliable and valid systems and in several instances makes reference to minimum group size for both reporting and comparison purposes. Unfortunately, conventional wisdom with respect to reliability and sample size and the practicalities associated with public schools forbid a simple application of a minimum n-size if both reliable and valid interpretations are being sought.

Much national debate on this issue has ensued. Within the debate, it is suggested that a minimum n-size to "guarantee" some reasonable degree of reliability might be as low as 100 students or as high as 350 students. Clearly this sort of a rule would result in the exemption of large numbers of schools from standard statistical comparisons of their performance and would eliminate consideration of almost all subgroup comparisons in almost all schools. This clearly was not the intent of the law (leave no child behind). Moreover, the guarantee of reliability in this instance results in a lack of validity (e.g. schools are treated differently based solely on size differences).

There is legitimate reason to be concerned when sample sizes become very small. With very small samples, estimates from year to year are very unstable and observed shifts in performance, unrelated to school effectiveness, are likely to occur. So, a desirable balance is to be sought between too many and too few students.

One solution lies not in the establishment of a specific "n-size" but in the employment of statistical controls. The use of confidence intervals is one such approach. Confidence intervals can be used productively to rule against year to year instability created by factors extraneous to school effectiveness (e.g. sampling error). One benefit of this approach is that minimum n-size requirements can be set very low. Confidence intervals can be computed with sample sizes as small as 5.

At first glance, it seems that confidence intervals are useful in order to deal with small samples of students. But in reality, the decision to use them is less based on sample size and more based on desire for "error-free" estimates. Error always pervades our estimates but confidence intervals allow us to control for some of the known contributors of measurement error (i.e. cohort differences). Toward this end, we can use confidence intervals that provide varying degrees of confidence. The level of confidence is predetermined prior to conducting the statistical tests. In this way, the same degree of confidence can be achieved when our sample includes 100 students or when it includes 25.

There has been significant outcry regarding the minimum n-size issue and many seem reluctant to rely on statistical approaches to control "error". Because of this, Nevada has chosen to use a hybrid approach in which it employs confidence intervals but does not conduct statistical tests on subgroups within schools with fewer than 25 students in the aggregate (e.g. summed across grades 3, 4, & 5).

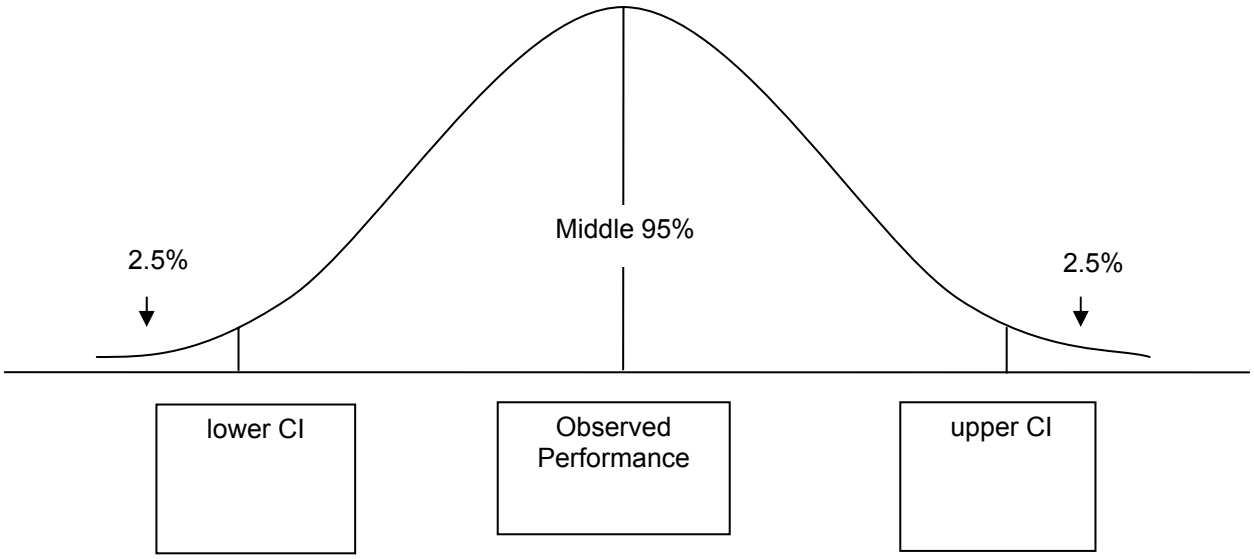
How do confidence intervals work?

Confidence intervals and their application are built on the basic measurement assumption that all measurements contain random error. In other words, an observed performance on a test is equal to "true" performance plus the effect of random "error". Examples of random error might include a dog barking outside the window of the testing room, no air conditioning, being sick, a test form that was positively biased in terms of the content most recently studied, or lucky guessing.

This means that for any given test administration, a student's observed score is as equally likely to be an overestimation as an underestimation of the student's "true" level of achievement. Sometimes our test scores suggest we are more knowledgeable than we actually are and at other times test scores suggest we are less knowledgeable than we actually are. Confidence intervals allow us to specify the "limits" within which true performance may fall. They allow us to set the upper and lower limits of performance estimation. For example, if a student scored a 50 on a test, confidence intervals allow us to know the likelihood that the student's true score is between 40 and 60.

The breadth or width of the "limits" depends on how much confidence we desire. Greater confidence results in a greater interval width. For example, if we score 50 and we are not too concerned about the accuracy of our judgment, we may expect that the true score lies between 45 and 55. If we are more concerned and desire more confidence, we may use 40 and 60. Intuitively, we can be more confident that the "true score lies between 40 and 60 than between 45 and 55 (e.g. what if the true score was 43?).

Using a graph of a normal distribution (e.g. bell-shaped curve), it is easy to see the relationship. We can look at the normal distribution of scores around the observed score and specify the amount of confidence being sought. In this example, a two-tailed 95% limit is illustrated.

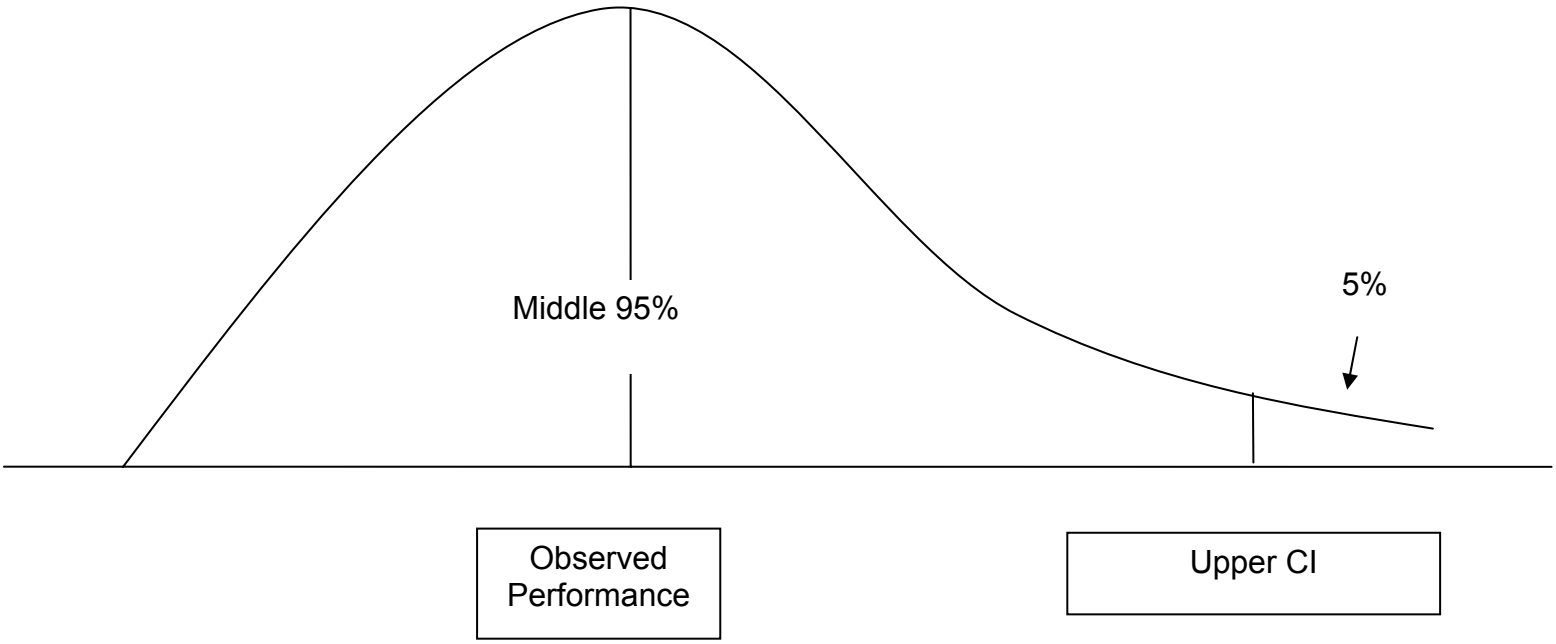


The graph nicely depicts how the observed score may underestimate performance or over emphasize performance. By organizing decisions using a two-by-two classification table, we can see the sorts of errors that can be made by using observed performance without considering measurement error.

	High Test Score	**Low Test Score**
High Achiever	Correct	*False Negative*
Low Achiever	*False Positive*	Correct

The table depicts the classification of high and low achievers based on test performance. From it we see that some high achievers score high on the test and others score low. When they score high we correctly classify them and when they score low we incorrectly classify them. This error is referred to as a false negative. By contrast, a low achiever may score high or low. A low achiever who scores high will be incorrectly classified as high achieving. This is referred to as a false positive.

There is always some likelihood that both errors can occur but attempts can be made to control for one or the other. When one is controlled there is greater likelihood that the other will be committed. Which we control for is a matter of choice and is often predicated on the consequences of the decision. For example, if we give accolades to high achievers and reward their achievement, we might want to be careful to be sure that the accolades are deserved and so we control against false positives. By contrast, if we sanction or punish low performance, we may want to be careful that the sanction is deserved and so we control against false negatives. Given the consequences associated with NCLB, the choice has been made to control against false negatives. In practice this means that we focus our attention on the upper limit of observed performance when we make comparisons.



By doing this we have made the a priori choice to ward against false negative classifications. By making this choice we must accept more false positive errors. This is a tough choice but prudent given how the AYP classifications are used. Moreover, the application of this sort of control can be done with virtually any sample size effectively evening the playing field.

It is understood that the use of statistical methodology can be a challenge. However, unless we accept the infallibility of "measurement", it is obligatory to account for unreliability when we make important decisions that ultimately impact the lives of students and educators. A conceptual understanding of measurement error prompted by graphic illustrations also

underscores how a narrow focus on sample size misses the point. Even with large samples, errors can be made. The point is that regardless of sample size, the use of confidence intervals better allows us to achieve reliable or consistent decision making.

AYP Indicators : Achievement, Participation, Other

Taking each AYP indicator in turn beginning with the achievement indicators and ending with the other indicators, we demonstrate the basic calculation for each. We then explore how each is used or compared within the AYP determination process.

Indicator Calculations

Achievement Indicators

Achievement Indicator 1: Status / PAC

The PAC (Percent at/above Cut) for a school is the primary achievement indicator for determining AYP. It is derived by simply dividing the number of students who demonstrate proficiency or performance indicative of meeting or exceeding standard by the total number of students.

$$\frac{\text{Number of meets or exceeds standard students}}{\text{All students}} \times 100$$

For example, if 1000 students were in the enrolled population, and 389 scored at or above the meets standard cut score, the PAC would equal 389/1000 x 100 or 38.9%. Note that in making comparisons, the NDE rounds percentages to the second decimal place. So if the total population hade been 900, PAC would equal 389/900 x 100 or 43.222222222222… or 43.22%.

Achievement Indicator 2: Safe Harbor / Relative Growth

Safe Harbor or, more appropriately termed, relative growth refers to the percent reduction in the percentage of non-proficient students. It relies heavily on the calculation of PAC but requires the PAC to be computed for adjacent years and the calculation of the difference in PAC from the previous year to the current year. So, using the same definition above for calculating PAC, relative growth is equal to:

$$1 - [\ (100 - PAC_{\text{current year}}) / (100 - PAC_{\text{previous year}})\ ]$$

For example, lets assume that in the current year the PAC was 30.00% and in the previous year it was 25.00%. Relative growth would be equal to 1 – [(100 – 30.00) / (100 – 25.00)] or 1 – (70.00 / 75.00) or 1 - .9333 or .0667. This means that the percentage of non proficient students was reduced by 6.67%.

Participation Indicator

Participation rate is a seemingly straightforward indicator. The mechanical definition for it is simple. It is equal to the number of students who took the test divided by the total number of enrolled students.

$$\frac{\text{Number of students who took the test}}{\text{All students}} \times 100 \text{ (rounded to second decimal place)}$$

For example, if 950 students took the test and 1000 students were enrolled, participation rate would be equal to 950/1000 x 100 or 95%.

The difficulty in determining participation rate is not in calculating the rate but in determining what constitutes "participation". At this juncture, NDE considers any student who has made a legitimate attempt at taking a test or one of several subtests as having participated. The only way for the NDE to determine this is by observing at least a single valid response to a test.

Other Indicators

Other Indicator 1: Average Daily Attendance (ADA)

Average Daily Attendance is also an intuitive and relatively easy indicator to calculate, assuming the necessary data elements are available. Considering and collapsing across all students in the school or subgroup within the school we divide the sum of positive attendance days by the sum of positive attendance days plus days absent:

$$\frac{\text{Sum of days present during the school year}}{\text{(sum of days presents) + (sum of days absent)}} \times 100 \text{ (rounded to second decimal place)}$$

To calculate this figure, we must know for each student the number of days they were considered in attendance and the number of days they were absent. For the 2002-03 school year, the NDE calculated these figures using information input into the Student Management and Automated Record Transfer System (SMART) by school districts. Because of the schedule of uploading information into the SMART system, the NDE had to use a truncated period of time to estimate ADA. The NDE used attendance information from the beginning of the school year until the beginning of December. This approach may have resulted in a slight overestimate of ADA among schools and subgroups within schools. For the 2003-04 school year, the NDE plan is to consider average daily attendance from the beginning of the school year through the first 100 days of instruction. Data to complete these school level analyses will be furnished by school districts and based on data pulled directly from their student information systems.

Other Indicator 2: Graduation Rate

Graduation rate is calculated to represent the school as a whole or any of its subgroups. Graduation rate is relatively intuitive indicator but quite complex in its calculation. It is also dependent on the availability of several pieces of information.

Graduation rate involves the estimate of graduation among a cohort of students. Although longer time periods can be considered, in its initial calculation graduation rate assumes the traditional 4-year high school schedule. Information must be collected on a cohort of students throughout those four years to generate the rate. This includes annual dropout rate figures, and figures pertaining to several separate completion options (e.g. standard diplomas, adjusted diplomas, certificates of attendance, GED recipients). Once collected the basic formula involves the division of students receiving standard diplomas by a combination of all completion possibilities and the numbers of students dropping out of school in 9^{th}, 10^{th}, 11^{th}, and 12^{th} grades. The calculation of graduation rate for the 2001-02 school year is as follows:

Number of standard diploma recipients (01-02)
---------- X 100 (rounded to second decimal place)
Number of standard diploma recipients (01-02) +
Number of adjusted diploma recipients (01-02) +
Number of certificate of attendance recipients (01-02) +
Number of GED recipients (01-02) +
Number of 9^{th} grade dropouts (98-99) +
Number of 10^{th} grade dropouts (99-00) +
Number of 11^{th} grade dropouts (00-01) +
Number of 12^{th} grade dropouts (01-02)

As is clear from the formula, much information with regard to a cohort of students is needed to calculate graduation rate. Most of the information is currently available for the calculation of graduation rate for whole schools and for race/ethnicity subgroups. At this juncture, historical data has not been collected on dropouts and some forms of completion for students with disabilities, students with limited English proficiency, and for economically disadvantaged students. Organized collections with respect to GED recipients are now under way as a consequence of the federal requirements.

Finally, two substantive differences exist between graduation rate and the other AYP indicators. First, the need for completion information makes it impossible to calculate the indicator for the "current" school year. In other words, for the 2002-03 AYP determinations graduation rates reflecting the graduating class of 2001-02 had to be used. Second, the graduation rate indicator collapses information across a four-year time span while the other indicators rely primarily on a single year of information. This means that change with respect to graduation rate is likely to take more time to observe.

Indicator Comparisons

Achievement Indicators

Status/PAC Comparisons:

As part of the AYP determination for a school, the PAC rate for the whole school and each of its identifiable subgroups must be compared against a statewide annual measurable objective or PAC target rate. As noted above, the PAC comparisons are made separately in ELA and math using only students enrolled within the school for a full-academic year or longer.

Although different PAC targets exist in ELA and math and comparisons must be made separately in each subject area, the same PAC goal must be used to judge whole school and subgroup performance. Moreover, subject area PAC goals are different for different clusters of schools. In Nevada, with few exceptions, schools are clustered in 3 ways: elementary schools, middle schools, and high schools.

In requiring the establishment of PAC targets or annual measurable objectives, the federal government required that the initial targets be based on baseline estimates of PAC established from assessments administered during the 2001-02 school year. Additionally, in establishing targets for the 2002-03 school year and beyond, states had to build targets to ensure that 100% of students would be proficient by the end of the 2013-14 school year. States were given the option to hold rates constant over predetermined time periods. However, changes in annual targets had to be equal distant.

States were provided two methods to establish baseline annual measurable objectives. However, states were required to use the method that yields the higher initial PAC rate. This resulted in the use of what might be termed the school percentile method.

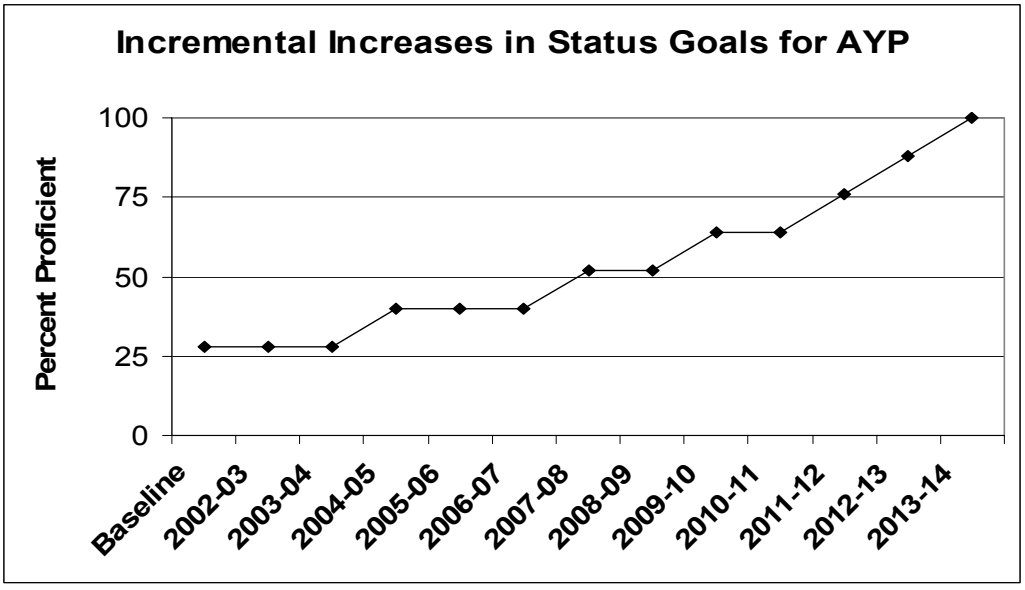
School enrollment percentile method

School Name	Percent Proficient	Enrollment	Cumulative Enrollment Percentage
School A	**2%**	**127**	**0.4%**
School B	**3%**	**28**	**0.5%**
School C	**6%**	**56**	**0.7%**
School D	**7%**	**380**	**1.97%**
School E	**10%**	**150**	**2.5%**
School X	**40%**	**281**	**20%**
State Total		**30000**	**100%**

The school percentile method involves the ranking of schools in terms of subject area PAC from lowest to highest. In addition, the enrollment for each school is indicated. The PAC rate for the school at the 20th percentile of cumulative state enrollment is chosen as the baseline rate for the given subject. The table directly above illustrates this method. Based on it, the PAC rate associated with school X would be used for baseline. Using this method, the following table includes the baseline estimates used for the 2002-03 school year.

	ELA	MATH
Elementary	30.0%	36.0%
Middle School	37.0%	32.0%
High School	73.5%	42.8%

When submitting its AYP plan to the federal government, the state was required to describe its method for establishing baseline and its estimated targets for each school year leading up to the 2013-14 school year. The state plans to use a tiered approach in which it will hold constant annual goals for as long as is allowed by the federal government while making associated equal distant target increases. The graph below depicts the state's tiered approach and the following table includes the actual estimated target amounts.



The annual goals depicted in the graph are for elementary ELA but the scheduled tiered increases reflect Nevada's plan for each subject area at each grad span level. For the 2002-03 and 2003-04 comparisons, annual targets were established by holding baseline estimates constant. The first increase, required by the federal government occurs for the 2004-05 school year. That goal is then held constant for two additional years followed by an increase for the 2007-08 school year.

Given the "equal distant" requirement, the target can only be held constant for one additional year before an increase in 2009-10. That target is again held for one additional year followed by increases in each of the last 3 years of the NCLB 12-year accountability time span. As noted in the graph, there are 6 increases in the 12 year period. Each of these increases must be equal. For any subject at any grade cluster, the 12-year plan can be established by subtracting the baseline PAC rate by the 100% goal and dividing that difference by 6. The resulting dividend is the necessary increase at each of the 6 time intervals.

Estimated Annual Measurable Objectives Through 2013-14 School Year

School year	Elementary School		Middle School		High School	
	ELA	Math	ELA	Math	ELA	Math
Baseline, 2002-03, 2003-04	27.5%	34.5%	37%	32%	73.5	42.8
2004-05, 2005-06, 2006-07	39.6%	45.4%	47.5%	43.3%	77.9%	52.3%
2007-08, 2008-09	51.7%	56.3%	58%	54.6%	82.3%	61.8%
2009-10, 2010-11	63.8%	67.2%	68.5%	65.9%	86.7%	71.3%
2011-12	75.9%	78.1%	79%	77.2%	91.1%	80.8%
2012-13	88%	89%	89.5%	88.5%	95.5%	90.3%
2013-14	100%	100%	100%	100%	100%	100%

The title of the table indicates that annual objectives are estimated. This is because of the following logic. As noted, the federal government required that states immediately implement the new accountability systems relying on what assessments were available during the 2001-02 school year. As noted above also, states are required to implement annual assessments in grades 3 through 8 and at the high school level by the 2005-06 school year. The implication of this in Nevada is that the assessment system used to establish baseline in the 2001-02 school year will undergo significant change until full implementation occurs in 2005-06. It is reasonable to expect some revision after that as well.

To address this inconsistency, the Nevada plan to the federal government indicated that as significant revisions occurred to the state assessment system, revisions to annual AYP targets could result. In no case can changes result in anything else than the 100% expectation in 2013-14 but the annual targets prior to that year could change.

The first known change to the annual targets occurs in for the 2003-04 school year. During 2003-04, criterion-referenced tests are being fully implemented at grade 8 and will result in the removal of the 7th grade norm-referenced test from AYP consideration. Similarly, the 4th grade NRT has been removed from AYP consideration and was replaced by the 4th grade writing test. Because baselines were so heavily dependent on the inclusion of NRTs, it has been judged prudent to reestablish school targets if the data supports a change. Based on analyses undertaken following the receipt of the 2003-04 data files, slight adjustments have been made at the elementary level and no adjustments were required at the middle school level (the above table reflects the elementary level changes). A second major revision effecting both elementary and middle school targets is expected to occur in 2005-06 or 2006-07 as the full implementation of grade 3 through 8 tests occurs.

The significantly higher PAC rates indicated at high school are worth mentioning. The basic difference between high school and the other school levels is in how the assessments feeding into the AYP calculations have been considered.

The federal regulations supporting the NCLB assessment prescriptions require that for AYP purposes states only use students' first testing opportunity. In other words, if a state administers a 3rd grade test more than once in grade 3, it is required to use the first administration of that tests to students in determining AYP.

This requirement created a difficulty for states such as Nevada that use high school exit examinations and have chosen to include those in the AYP system as opposed to creating a separate set of accountability measures. The "exit" examinations in Nevada are used to measure proficiency as students are exiting high school. However, in the service of fairness, the state begins providing opportunities for students to take these examinations as early as grade 10. It would be inappropriate to rely solely on 10th grade performance when we know that the test is designed to measure content that students might not be exposed to until after that time period.

Because of this situation, the Nevada plan requested the use of a cumulative exit examination pass rate in establishing annual PACs. The request was to include up to 5 test opportunities through April of a student's 12th grade year of instruction. That request was denied but the federal government allowed the inclusion of the first two student attempts. Still, the allowance of two attempts is different in nature to the allowance of a single attempt in middle school and elementary school. This difference is why the baseline PAC rates in high school are larger.

A 3rd revision to the annual targets could occur at high school. Nevada did ask that the federal government reconsider the request for a broader cumulative rate once an efficient tracking mechanism of student performance was in place. Nevada expects to request a revision beginning with the 2004-05 school year.

Returning to a discussion of status or PAC comparisons, once the annual targets have been set, each year the PAC for the whole school and each subgroup is compared against the annual target. For example, in elementary school the annual ELA PAC annual measurable objective in 2004-05 is 39.6%. To meet the status requirement the whole school and each subgroup would have to have an ELA PAC rate of at least 39.6%.

To buttress the reliability of the status comparisons, the standard error of the proportion is used to adjust observed PAC scores before a comparison against the state target was made. This is an important step allowing for a predetermined degree of confidence in the status comparisons.

The standard error of the proportion accounts for sampling error that might be associated with unique characteristics of a particular cohort of students. The formula for the adjustment is relatively simple to apply. We multiple the proportion of proficient students with the proportion of non-proficient students and divide that quotient by the number of students in the sample. Then we take the square root of that dividend.

$$\sigma_P = \sqrt{PQ / N}$$

For example, if 120 students in the school took the ELA tests and 36 scored at or above proficient (30% PAC) we would multiply 30% by 70% and divide that by 120. This equals .0027. The square root of .0027 is .0459. The result is a single standard error. The single error adjustment provides more certainty than relying on the observed score alone but the level of certainty might be considered low using conventional applications for statistical tests. To ensure at least a 95% certainty or "confidence", we can multiple the standard error by 1.645. Applying this rule to our standard error results in an adjustment of .0459 x 1.645 or .0754.

We would add the 7.54% correction to the observed PAC of 30.00%. Thirty plus 7.54 is equal to 37.54. The adjusted PAC of 37.54% is what is compared against the state annual measurable objective (AMO) of 39.6% and in this case the school would be deemed to have not met the status achievement comparison.

It is important to note that the motivation to use the adjusted score is not to give schools an advantage or to lower expectations. The use of one-tailed confidence intervals in this instance helps to better ensure the reliability of the system by warding against false negative classifications. Having a reliable system is a federal requirement. By making the adjustment as noted above, we have a relatively high degree of confidence that the school's "true" level of performance is below the state target.

Safe Harbor/Relative Growth Comparisons:

For schools or subgroups that fail to meet the status goal, they can still be deemed as making AYP if the percentage of non-proficient students is reduced by 10% or more and the group in question meets the criteria of the other indicator analysis. Therefore, the safe harbor/ relative growth comparison is a conjunctive analysis in which **both** a 10% reduction in the percentage of non proficient students **and** performance exceeding the target on the other indicator is observed (see the next section for information on other indicator comparisons). This analysis can be viewed as a secondary analysis provided to educational agencies that do not meet the PAC annual measurable objective.

For example, if in year 1 the percentage of non-proficient students was 63.5%, a 10% reduction would be equal to 63.5 multiplied by .10 or 6.35%. More concretely, this would mean that the percentage of proficient students in year 2 would need to be 6.35 percentage points higher than the percentage of proficient students in year 1, for a 10% reduction in non proficient students to be observed.

As with status/PAC, only students enrolled for a full-academic year are included in this comparison.

For this analysis, the NON-PAC rate from the current year is divided by the previous year NON-PAC rate. The resulting ratio is then subtracted from 1 to obtain the observed percent reduction in the percentage of non-proficient students last year. The change is compared to the 10% change threshold. As with status comparisons, the state uses confidence intervals to ensure more reliable decisions. There is a difference in the appropriate error estimate used to establish the

interval. Because the PAC rate from two separate administrations is considered, control over sampling error from both administrations must be achieved.

Toward this end, the standard error of the difference in proportions is used. The application of the formula is very similar to the application of the standard error of the proportion. We must take the square root of the sum of the variance estimate for each separate administration.

$$\sigma_{P1-P2} = \frac{\sqrt{\sigma_{P1}^2 + \sigma^2_{P2}}}{\text{NON-PAC}_{\text{Previous Year}}}$$

To help explain the safe harbor calculation, we provide an example. In the current year 36 of 120 students or 30.00% were proficient whereas in the previous year 25 of 100 or 25.00% of the students were proficient. We get the change in NON-PAC rate by dividing the current year NON-PAC (1-PAC) rate by the previous year NON-PAC rate and subtracting the ratio from one. This works out as:

1 - [(100 - 30.00) / (100 - 25.00)] = 1 – (70.00 / 75.00) = 1 - .9333 = .0667.

It is this difference that must be adjusted to account for sampling error.

The adjustment is derived using 5 steps. First, multiply the proportion of proficient students by the proportion of non-proficient students and divide that result by the number of participating students to obtain the variance estimates. This must be done for both years in question. Second, sum the variance estimates. Third, take the square root of the summed variance estimates. Step three yields the standard error of the difference in proportion. Fourth, multiple the standard error by the appropriate z-value to establish the a priori determined confidence limit. The federal government required that Nevada cap the confidence associated with Safe Harbor comparisons at .75. A z-score of .675 establishes this limit. Finally, divide the resulting confidence interval by the percentage of non-proficient students in year one to convert the confidence interval for the difference in proportions to the confidence interval for the percent difference in proportions.

So for our working example:

Standard error = Square root of { [(.3000 *.7000) / 120)] + [(.2500 * .7500)/100)] } = .0602
Z-score transformation to .75 limit = .0602 * .675 = .0410.
Conversion to CI for percent difference in proportions = .0410 / .75 = .0547

To adjust our observed difference we add the confidence interval to the observed difference or .0667 + .0547. This equals .1214 or 12.14%. This number would be compared to the 10% change requirement and in this case we would judge the school to have met the safe harbor/relative growth comparison.

Participation Indicators

Test Participation Rate Comparison:

The comparison of this indicator is relatively strait forward. As a reminder, all students enrolled at the time of testing must be included in the participation rate calculation. Participation rate must be calculated separately for ELA and math. No correction for measurement error is applied to this comparison.

The observed participation rate is compared against the 95% participation rate for the whole school and for each of its subgroups. This is a simple comparison but 95% is a rigorous standard. Moreover, if strictly applied for a school or subgroup within a school with 20 students, all but one student would have to participate to meet the criterion. If the school or subgroup had 19 or fewer students, all students would have to participate to meet the 95% criterion (e.g. 18/19 = 94.7%).

There are legitimate circumstances that can result in a student's failure to participate. In an attempt to be sensitive to this, a modified criterion has been established for schools or subgroups within schools that have fewer than 20 students. It is labeled the N-1 rule. Instead of using the 95% threshold in this instance, we apply a standard of N-1 with N being equal to the number of enrolled students at the time of testing. For instance, if a school had only 19 students, at least 18 of the students would have to participate. Likewise, for a school with 10 students, at least 9 would have to participate.

Other Indicators

Other Indicator Comparisons:

The application of these comparisons is relatively straight forward. At this juncture, no measurement error correction is used with these comparisons. As a reminder, all students enrolled at the school during the school year are included in this calculation. Additionally, only the school as a whole, not ethnic or special subgroups (i.e. IEP, LEP, FRL), is judged against the other indicator target as a stand alone AYP analysis. Subgroup performance on the other indicator is only considered for subgroups if a safe harbor analysis is needed because the subgroup did not meet the Status analysis target (i.e. AMO).

The first step was to establish state goals for these indicators. For ADA, Nevada statute requires a 90% student attendance rate. The state adopted this threshold to use as a goal and to judge each separate subgroup.

For graduation rate, no such standard existed. To explore alternatives, the Department of Education applied methods provided by NCLB to establish achievement baselines. Using this as reference, the State Board of Education has temporarily adopted a graduation requirement of 50%. The State Board of Education is expected to revise this standard after appropriate information is gathered.

For each indicator, a school or subgroup within a school can meet the AYP requirement in two ways. First, if they perform at or above the threshold they have met the AYP requirement. Second, if they are below the threshold but have made some positive gain in comparison to the previous year, they are deemed to have met the criterion.

For example, if an elementary school's ADA is at or above 90% they have met that AYP criterion or if it is below 90% but it is greater than its ADA in the previous school, it is still okay.

Section Summary

In this section, outlined were the key variables to be considered in making AYP determinations. Measures used to increase the reliability of our decisions were considered as were the operational definitions and calculation steps for the AYP indicators.

Putting all of this together allows the state to conduct school and district level comparisons and to make preliminary AYP determinations. The next section outlines the transition from preliminary determinations to final determinations.

School and School District AYP Classifications and Annual Achievement Designations

As discussed, schools and school districts are judged annually and classified as having met or having failed to meet AYP. Based on AYP classifications, schools and school districts receive achievement designations. Designations include an identification of schools that are in need of improvement, high achieving, exemplary in achievement, and that are most improved. Schools that receive none of these distinctions are considered to be adequate in regards to achievement.

NCLB gives the ultimate authority for making school level classifications to the school district. It gives the state the authority to make school designations and school district classifications and designations. Senate Bill 1 mimicked the NCLB language providing authority in a similar fashion. However, it requires the state to make all preliminary classifications.

Using the steps outlined above, the state conducts an analysis for every school and school district. We discuss school and school district classifications first followed by school and school district achievement designations.

<u>School AYP Classifications</u>

Based on data collected from test score sheets and limited information provided from other data sources (i.e. other indicator performance), the state conducts the AYP comparisons. This results in a profile for each school that summarizes the states findings.

The state issues the preliminary information to local school districts for distribution to their schools. The information is considered preliminary until both schools and school districts have ample opportunity to review the analyses and determinations.

Although legally schools are given an opportunity to appeal achievement designations, schools are given an opportunity to appeal their AYP classifications to the school district. Senate Bill 1 requires that school districts consult with the NDE in judging appeals and prior to making final determinations. School districts must furnish the state with comprehensive support materials to assist the state in consulting on school level appeals (i.e. student data files and output from reanalysis of data). Again, the final school determinations are made by school districts and not by the state. School districts must provide to the state a final listing of schools, by AYP content area that are classified as not having met AYP.

The opportunity for appeal is given in the interest of making valid decisions. For example, the school may believe that the state analysis is in error for a variety of reasons or may believe that other evidence could be introduced that paints a different picture of school performance.

For the 2002-03 school year, the grounds for appeal were applied liberally due in part to the transition to the new accountability system but mostly because of the delay in both federal regulation and state legislation providing the authority to implement the new system. The NDE and local school districts have worked jointly to specify grounds for appeals for the 2003-04 school year and beyond. Although it is the ultimate authority of each school district to classify its own schools, there is interest in seeing an equitable and common application of the process across school districts. Appeals may be granted after review if student performance was adversely affected by extraordinary and unavoidable circumstances during testing, if significant coding errors impact the AYP analysis[2], if additional statistical analyses conducted by the school or school district identify errors in the state calculations, or if other significant factors produce statistical or substantive explanations for school performance.

School District Classifications→ At the same time that the state issues school level preliminary classifications, it must issue preliminary school district classifications. The school districts then have an opportunity to appeal their classifications directly to the NDE.

The NDE employs a nearly identical list of appeal grounds in judging school district appeals. When considering school district appeals, the NDE requires that school districts provide backup materials supporting their appeals.

The need for backup materials is due to the limited amount of information available at the state level for use in the AYP determination process. As noted previously, with few exceptions, the state relies on the student level information provided on test score sheets. When schools appeal classifications to school districts based on coding issues, for example, the school district has at its disposal student level records that can be used to verify findings. Currently, no such store of information is available at the state level to verify district appeals. This places the onus on school districts to clearly support the appeals they present.

[2] Coding errors may occur for a variety of reasons. Those judged to be legitimate errors, and not errors due to negligence or errors that are repetitive from past years, may result in a reanalysis of AYP. Reanalysis is the responsibility of the school and school district.

After the close of the district appeal window, the NDE releases its final school district classifications by category. Below is a table outlining the general timeline from the CRT test window to final AYP determinations to implementation of school/district improvement plans.

March 15 to April 15	CRT Testing window
April 22nd	Score materials to test vendor
May 20th	Assessment reports to school districts
June 15th	NDE issues preliminary school and school district AYP classifications
July 1st to July 31st	Designation appeal window
August 1st	Final designations made
August 15th	Dissemination of District Report Card
September 1st	Dissemination of State Report Card
November 1st	School improvement plans due date
January 1st	Implementation of school improvement plan

Achievement Designations

In Need of Improvement (INOI) Designations

Once success or failure on the annual AYP judgment has been formally determined, the tracking of schools ensues. As noted earlier, schools must be judged separately in ELA and Math as a requirement of NCLB. Following this logic, Nevada tracks schools by subject areas and the other indicators separately. This becomes a significant as we consider achievement designations.

To be designated as INOI a school must fail AYP in two consecutive years. In Nevada, this means that the failure must occur in the same subject area or relative to the other indicator in consecutive years. To assist in the designation process, schools are placed on what has been termed "watch" lists. The lists are distinguished by which areas schools are being watched for.

Single Year Classification Table

ELA	Math	Other Indicator	AYP Decision	Improvement Classification
Pass	Pass	Pass	Meets AYP	Adequate
Fail	Pass	Pass	Fails AYP	Watch (ELA)
Pass	Fail	Pass	Fails AYP	Watch (Math)
Pass	Pass	Fail	Fails AYP	Watch (OI)
Fail	Fail	Pass	Fails AYP	Watch (ELA & Math)
Fail	Pass	Fail	Fails AYP	Watch (ELA & OI)
Pass	Fail	Fail	Fails AYP	Watch (Math & OI)
Fail	Fail	Fail	Fails AYP	Watch (All 3)

Two Year Designation Table

Year 1 AYP Classification	Year 2 AYP Classification	2-Year Achievement Designation*
Meets AYP	Meets AYP	Adequate
Fails AYP	Meets AYP	Adequate
Meets AYP	Fails AYP (any combo)	Watch ("combo" elements)
Fails AYP (ELA)	Fails AYP (Math)	Watch (math)
Fails AYP (Math)	Fails AYP (ELA)	Watch (ELA)
Fails AYP (ELA)	Fails AYP (ELA)	In Need of Improvement
Fails AYP (Math)	Fails AYP (Math)	In Need of Improvement
Fails AYP (Both)	Fails AYP (either)	In Need of Improvement
Fails AYP (OI)	Meets AYP	Adequate
Meets AYP	Fails AYP (OI)	Watch (OI)
Fails AYP (OI)	Fails AYP (OI)	In Need of Improvement

* Table does not include an exhaustive set of designation possibilities and should be considered illustrative

For example, a failure in year 1 in ELA followed by a failure in ELA in year 2 results in an INOI designation. However, failure in ELA in year 1 followed by a failure in Math but passage in ELA in year 2 would not result in an INOI designation. Instead, the school would move from being on watch for ELA to on watch for Math. A failure in Math in year 3 would move the school into INOI status.

Once identified as INOI, a school must successfully pass the associated AYP criteria for two consecutive years to have the label removed. Following the same logic as noted above, a school designated as INOI because of Math failure would need to pass the Math requirements for two consecutive years to have the label removed. A failure in ELA in either of the two years, but not both consecutively, would not affect the school's INOI status, but the school would be watched relative to this failure.

Exemplary and High Achievement Designations

While the AYP classification system is designed to identify schools that require technical assistance and support it is also designed to identify schools that are exceeding state expectations for performance. This is a key attribute of the model if it identifies schools that can serve as models for "like" lower performing schools.

As required by state statute, the State Board of Education must develop criteria to be used in designating schools as high achieving and exemplary. The following is the basic set of criteria that must be met for a school/school district to earn such a distinction.

The school or school district must have made AYP in the current school year and cannot currently be designated as "In Need of Improvement". In addition to meeting these requirements:

To be designated as High Achieving:

1) The percentage of students in each identifiable subgroup that score at or above the level of "meets standard" in each subject area must be significantly greater than the annual measurable objective or PAC requirement; ***or***
2) For the school as a whole (not subgroups), the reduction in the percentage of non-proficient students (students scoring at or above meets standard) must decrease by significantly more than 10% from the previous school year.

To be designated as Exemplary:

1) The percentage of students in each identifiable subgroup that score at or above the level of "meets standard" in each subject area must be significantly greater than the annual measurable objective or PAC requirement; ***and***
2) For the school as a whole (not subgroups), the reduction in the percentage of non-proficient students (students scoring at or above meets standard) must decrease by significantly more than 10% from the previous school year.

For the PAC comparisons, a significant difference for the 2003-04 school year is based on a one-tailed 95% confidence interval. This means that the lower tail of the observed PAC for the school and subgroup when relevant must be greater than the annual measurable objective or PAC in the content domain. For the reduction in non-proficiency comparison, a significant difference for the 2003-04 school year is based on a one-tailed 75% confidence interval meaning that the lower tail of the observed decrease for the whole school must be greater than 10%.

For example, the PAC objective in ELA at the elementary level is 30%. Using the formula provided previously for calculating the standard error of the proportion and the z-score transformation, for a school of 25 students there observed PAC in ELA would have to be 46.5 or higher to be judged as significantly greater than the annual measurable objective of 30%. Likewise, using the formula for the standard error of the difference in proportions and assuming that the size of the school was not different in the previous year, the PAC rate for the school in the previous year would have had to have been 28.7 or lower for the reduction in non-proficiency to be judged as significant.

To aid in making significance judgments, the Department of Education will publish tools available at its website that enables easy analysis of PAC and Safe Harbor comparisons. The tool ("AYP Calculator") will assist schools and school districts in recalculating AYP for the purpose of AYP appeals but can also be used to target the level of performance necessary to be considered high achieving or exemplary.

Most Improved Designations

All school/school district designations discussed thus far are based on the distinction between proficient and non-proficient students or students that either meet or fail to meet standard. As required by the NCLB Act, performance on the state tests used to calculate AYP is reported

relative to 4 achievement levels, not just two. In addition to the "meets standard" level of achievement, students can be classified as "exceeding standard", "approaching standard", or as developing/emergent.

In keeping with the intent of the NCLB Act, it is important that students be challenged to progress regardless of their current level of achievement. For example, a student that demonstrates that she has met the standard ought to be challenged to exceed expectations. Likewise, a student that demonstrates the lowest level of achievement (developing/emergent) should be encouraged and rewarded for significant progress even if his progress is still below the meets standard cut point but greater than where he began.

Toward this end, a school or school district that demonstrates significant movement of students into higher achievement levels in both English Language Arts and Mathematics may be recognized as demonstrating significant improvement. For the 2003-04 school year, significant improvement is based on a reduction in the percentage of students in the lowest achievement level of at least 7.5% and an increase in the percentage of students in the highest achievement level of at least 7.5%. In addition, there cannot be an observed decrease in the percentage of proficient students for any identifiable disaggregated group.

Reporting of Annual AYP Judgments

It is anticipated that the state will report the annual judgments in multiple ways. First, it is likely that press releases will be issued that list schools being placed on watch and schools classified as INOI. Reasons for classification are summarized in these releases. The Department does not intent to release preliminary findings but may release information after school districts have made final designations.

With more detail, lists of schools by subject area/other indicator failure are provided on the department web page. It is anticipated that the lists will allow direct links to AYP profile information supporting the classification. Toward this end, schools not failing will also be listed so that their profiles can be accessed electronically.

Finally, Senate Bill 1 requires that information relative to both AYP and INOI classifications be published annually. This includes a listing in the annual accountability reports of those schools failing AYP on an annual basis. It is important to note that this reporting requirement is state specific and not required by NCLB. Additionally, schools identified as INOI must be listed along with an indication of the number of consecutive years in which they have had that label.

Section Summary

Discussed was the general approach taken to formally classify schools and report findings. Review of similar documentation provided by school districts ought to be reviewed as the contents of this report are from the perspective of the state only. As noted, school districts share significant authority in this process and have undoubtedly prepared their own guidance toward this end.

The preceding provided general information, steps and processes that are applied to all Nevada public schools when practical (private schools are not governed by the described NCLB and SB1 accountability requirements). There are a variety of situations and special circumstances that prohibit the "easy" application of these rules to certain schools. Some exceptional circumstances are described next.

Special Circumstances

Public schools in Nevada share many characteristics and the majority of schools share similar configurations. For example, most schools follow K-5, K-6, 6-8, 7-8, or 9-12 grade configurations. However, there are anomalous configurations such as 5-8 or K-12. Moreover, some schools serve only specialized subgroups of students or contain magnet programs that serve a specific group of students. These anomalies prohibit the application or make the application of the general AYP rules more difficult.

Small Schools

As noted above, a minimum n-size is not necessarily required to make statistical comparisons but a policy decision has been made to only compare results for subgroups if and when their aggregated total is at least 25. This creates a problem when the total school population, aggregated across testing grades, is less than 25. It is important to note that these schools must still be judged in some fashion. They cannot be exempted form the AYP determination process.

For the 2002-03 classifications, the typical AYP rules were applied to small schools but small schools were provided some extra flexibility in appealing classifications. The basic justification was that the statistical estimates are less stable as a result of very small sample sizes. In this instance, small schools could introduce local assessment data that might change the interpretation of overall school performance. Local assessments have to align to state standards in order to be used in this fashion.

For 2003-04 and beyond, the state intends to aggregate data for very small schools across years until the n = 25 threshold is achieved. In other words, following the same aggregation logic described earlier, a particular school's results from the 2002-03 school year will be combined with the 2003-04 results before making the 2003-04 determination. These aggregation steps can be used to combine data across two or three adjacent years if necessary.

This approach cannot solve all the issues associated with very small schools and so the state still intends to carefully judge appeals presented by small schools to school districts.

Anomalous Grade Configurations

There are two general anomalies associated with grade configurations that affect the application of AYP. The first is anomalous grade spans that bridge two or more levels of instructional programming (i.e. elementary & middle, middle & high, elementary through high). For example, some schools in the state encompass grades 5 through 8. Grade 5 is typically considered as part of the elementary program, while grade 8 is considered part of the middle school program. For these schools, data was collapsed across all grades in which assessments

were administered. In practice, K-8 schools were judged against elementary statewide AYP targets, grade 5-8 schools were judged using middle school statewide AYP targets, and K-12 schools were split using the elementary targets to judge performance among K-8 students and high school targets from students in grades 9-12.

The second grade configuration anomaly is the instance in which the school only serves students in grades in which no state tests are administered (i.e. grades k-2). In this instance, only "other indicator" information is considered in classifying the school. There has been consideration of "backtracking" student performance to K-2 schools from the schools they feed. Until an efficient tracking mechanism is available, this alternative is not practical.

Alternative Programs

NCLB makes no distinctions regarding school types and the application of AYP and instead reinforces the need to hold all publicly enrolled students, schools, school districts accountable under the AYP system without exception. Hence, the rules that apply to traditional schools in Nevada are applied to what have been termed alternative schools.

The majority of students enrolled in alternative schools/programs are there as a result of difficulties they have experienced in traditional settings. The achievement levels among these students are notably low and most of these schools serve students that are assumed to be at-risk for academic failure.

But there is no allowable exception. Alternative schools like all traditional schools are judged using the standard achievement, participation, and other indicators. As is true with traditional schools, school achievement is based only on students who have been enrolled in the school for a full academic year.

Schools or Special Programs for Students with Disabilities

Similar to alternative schools/programs, in Nevada there are a handful of public schools dedicated to serving students with disabilities. Some of the dedicated programs only serve students with rare and extremely debilitating disabilities. In many Nevada public schools, there are special programs within the school that serve a particular disabled population. These programs are at times referred to as magnet programs. Often they serve students that are not zoned for enrollment in that location.

As with alternative schools, no exceptions can be made for students enrolled in public schools/programs. The federal law requires that they be accounted for and that the school must be judged relative to their achievement etc…

There has been some consideration for applying the achievement scores earned by these students back to their "zoned" school. There has also been consideration of treating magnet programs within schools as a school unto themselves. Firm policy has not been developed at this juncture and until a formal change is made, these students as a subgroup are treated as all students enrolled in the school in which the assessments are administered.

Charter Schools

There is no distinction to be made between traditional public schools and public charter schools in Nevada in terms of the application of AYP at the school level. However, Senate Bill 1 passed after the federal review of the Nevada Accountability Work Plan prevents the aggregation of performance within charter schools up to the sponsoring school district level. Because the Nevada Accountability Plan did not address this state legal requirement, an amendment to the federal accountability plan will have to be made.

Once again, the same AYP rules apply to charter schools just like all other public schools. The performance of students enrolled in charter schools for a full academic year is used to judge charter schools. The performance of students in charter schools is aggregated to the state level. The state is responsible for and held accountable for all students enrolled in its public educational programs.

Correctional Programs

As is true for alternative programs, special programs, and charter schools, the federal government to date has not allowed a distinction to be made between traditional public schools and correctional facilities that provide educational programs. Therefore, at this juncture correctional programs are reviewed annually by applying the same AYP rules that are used with more traditional educational programs.

One of the chief concerns that has been raised regarding the students in these programs is that their typical stay in the facility is less than a full academic year. As is true with all other schools, only students that have been enrolled in the program for a full academic year are included in AYP analyses. However, the correctional programs are embedded within school districts and assuming that at least some of the students enrolled in the correctional programs have been enrolled in the school district for longer periods of time, the results for students may be aggregated up to the school district level even when their performance has not impacted the school level judgment.